CLASSIFICATION OF STATISTICAL SERIES.

By Edmund E. Day

Classification of statistical series is more than an exercise in applied logic; it is the beginning of statistical analysis. Statistical methods to be reliable must respect the character of the data in hand. Dependable scientific results rest as much upon knowledge of the conditions under which a given operation is desirable as upon skill in the performance of the operation finally selected. What is admirable practice for one body of material is unpardonable malpractice for another. Recognition of the fundamental differences between statistical series of different types is thus the first step in satisfactory analysis.

Classification is always shaped by the purpose for which classification is undertaken; corresponding to a large variety of purposes, a large number of different classifications may be evolved.* However, in the classification of statistical series, by far the most generally important purpose is the development of statistical method. What is needed is a classification which will materially strengthen and clarify the exposition of statistical theory.

No clear notion of such a classification of series is to be obtained from the standard treatises on statistical theory. Several of the most widely employed texts appear practically to ignore the matter.† Other works recognize by implication the essential differences between historical data and other types of statistical material, but get no further.‡ Professor Horace Secrist, in his recent *Introduction to Statistical Methods* (New York, 1917), divides statistical tables into three classes: (1) "those which express historical data"; (2) "those which

*Cf., F. Zizek, *Die Statistischen Mittelwerte* (Leipzig, 1908), ch. 1. Translated under title of "Statistical Averages," by W. M. Persons (New York, 1913).

†*E. g.*, G. U. Yule, *Introduction to the Theory of Statistics* (London, 1912), and C. J. West, *Introduction to Mathematical Statistics* (Columbus, 1918). The treatment of method in both of these texts is almost exclusively from the viewpoint of the frequency, or attributive, distribution.

‡*E. g.*, A. L. Bowley, *Elements of Statistics* (3d ed., London, 1907); W. I. King, *Elements of Statistical Method* (New York, 1914).

describe a situation or condition in cross-section"; (3) "those which express variable data of non-historical character."* But the logical basis of this classification is not indicated, and the distinction between the second and third classes is nowhere made clear.† Only among some of the Continental statisticians has the classification of series been carefully considered.‡ Unfortunately among these writers the full import of classification for the development of statistical theory has not been realized. In general, the classification of statistical series and the exposition of statistical theory have influenced each other but little.

A classification, serviceable for the purposes of statistical analysis, must be based upon the relation between statistical series and the original individual observations. It is a common-place of statistical discussion that even the simplest observations involve three elements: (1) a location in time; (2) a location in space; and (3) the existence (or non-existence), possibly the existence in varying degree, of specified characters or attributes. These basic elements are the raw stuff from which statistical series are made. Naturally paralleling the different observational elements are the different types of series. A series may be constructed: (1) of individual observations without significant time or space differences but exhibiting some attribute in varying degree; or (2) of individual observations which are alike except that they are differently located in space; or (3) of individual observations which differ only in their time limits. Thus the three fundamental types of series are: (1) attributive—commonly called frequency—distributions; (2) spatial distributions; (3) temporal distributions.

The value of this classification lies in its illumination of the problems of statistical analysis. The different phases of statistical method are of varying importance for the three types of series. The customary discussion of averages, for

*See pp. 142–145.

†Furthermore, against Professor Secrist's attempt, it may fairly be urged that it is less important to classify statistical tables than statistical series, since a single table may contain a variety of series.

‡*Cf.*, G. v. Mayr, *Statistik und Gesellschaftslehre* (2d ed., Tübingen, 1914), Vol. I, pp. 142–149; Al. Kaufmann, *Theorie und Methoden der Statistik* (Tübingen, 1913), pp. 487–488.

example, relates primarily to the attributive distribution. In the case of time series, the relative number is more directly significant than the average. Furthermore, time series are peculiar in that they require a separation of the different varieties of temporal variation: normal, or secular, trend; cyclical fluctuation; seasonal variation. Spatial distributions afford exceptional opportunities for the graphic method, which in the exhibition of geographic relationships possesses advantages over any other mode of analysis. It would be easy to multiply examples of the desirability of recognizing in the explanation of statistical methods the differences between the underlying types of statistical series. The fact that the more refined methods of statistical analysis have been developed in the study of attributive distributions, whereas most economic materials assume the form of time series, is further reason for the classification of series in the extension of statistical method to economic investigation. It is to be hoped that future texts on statistical theory will see the gain to be secured by classification, and will order their exposition of statistical methods with adequate regard for the essential differences between the fundamental types of statistical series.